Top 10 Words Per File

=== 106247newsML.txt ===

internet: 9, league: 6, site: 5, consumers: 4, said: 4, fraud: 4, may: 3, scam: 3, investors: 3, web: 3

=== 120600newsML.txt ===

mail: 7, said: 7, post: 7, office: 7, service: 6, reisner: 6, internet: 5, postal: 4, electronic: 4, security: 3

=== 120683newsML.txt ===

school: 3, study: 3, access: 2, districts: 2, city: 2, elementary: 1, students: 1, internet: 1, learned: 1, kids: 1

=== 136958newsML.txt ===

internet: 7, proposal: 7, said: 7, society: 6, committee: 6, names: 5, heath: 5, authority: 4, addresses: 3, august: 3

=== 137498newsML.txt ===

internet: 8, proposal: 7, society: 6, authority: 4, addresses: 3, committee: 3, names: 3, international: 3, said: 3, one: 3

=== 14014newsML.txt ===

trademark: 2, internet: 2, association: 2, group: 1, leading: 1, specialists: 1, plans: 1, release: 1, recommendations: 1, aimed: 1

=== 156814newsML.txt ===

online: 7, federal: 7, said: 6, information: 5, commission: 5, laws: 4, internet: 4, social: 4, company: 3, web: 3

=== 182596newsML.txt ===

futures: 9, commodity: 8, banks: 8, foreign: 7, court: 6, amendment: 6, would: 6, said: 6, currency: 6, cftc: 6

=== 186392newsML.txt ===

cftc: 8, foreign: 7, dunn: 7, court: 6, currency: 6, trading: 6, commodity: 5, case: 4, regulation: 4, said: 4

=== 193495newsML.txt ===

online: 7, internet: 6, poll: 5, people: 5, found: 4, said: 4, harris: 4, computers: 4, access: 4, hours: 4

=== 196805newsML.txt ===

encryption: 10, said: 6, key: 5, internet: 4, products: 4, new: 3, plan: 3, sales: 3, computer: 3, export: 3

=== 197734newsML.txt ===

online: 7, internet: 6, poll: 5, people: 5, found: 4, harris: 4, computers: 4, access: 4, hours: 4, year: 3

=== 206838newsML.txt ===

tax: 12, said: 8, report: 8, treasury: 5, new: 4, department: 4, internet: 3, government: 3, policy: 3, issues: 3

=== 231479newsML.txt ===

policy: 9, export: 8, said: 8, products: 8, alliance: 7, software: 6, key: 6, rules: 5, encryption: 5, recovery: 5

=== 233150newsML.txt ===

credit: 19, union: 10, unions: 9, rules: 9, members: 9, jackson: 8, said: 8, court: 7, new: 7, single: 6

=== 237175newsML.txt ===

said: 9, report: 8, internet: 7, magaziner: 7, draft: 6, policies: 4, commerce: 4, could: 3, provide: 3, government: 3

=== 249407newsML.txt ===

said: 7, export: 6, rules: 6, meeting: 6, products: 6, reinsch: 5, draft: 4, key: 4, clinton: 3, computer: 3

=== 2537newsML.txt ===

web: 9, security: 9, said: 9, site: 7, hackers: 7, department: 5, sites: 5, justice: 4, magazine: 4, internet: 3

=== 266038newsML.txt ===

banks: 10, said: 5, fed: 5, congress: 3, percent: 3, bank: 3, comptroller: 3, ludwig: 3, insurance: 3, funds: 3

=== 269995newsML.txt ===

banks: 12, said: 9, securities: 7, fed: 6, firms: 5, bank: 5, congress: 5, subsidiaries: 4, underwriting: 4, percent: 4

=== 269999newsML.txt ===

data: 6, discrimination: 6, fed: 6, regulation: 6, department: 4, said: 4, justice: 3, banks: 3, discourage: 3, lending: 3

=== 270046newsML.txt ===

internet: 11, names: 7, new: 7, committee: 6, sites: 5, international: 4, said: 4, addresses: 3, number: 3, domain: 3

=== 270084newsML.txt ===

internet: 12, names: 8, committee: 7, new: 7, said: 5, international: 4, addresses: 4, number: 3, domain: 3, society: 3

=== 270134newsML.txt ===

banks: 11, said: 11, securities: 10, fed: 9, act: 5, congress: 4, industry: 4, underwriting: 3, section: 3, percent: 3

=== 270346newsML.txt ===

banks: 9, securities: 7, fed: 6, firms: 5, bank: 5, said: 5, subsidiaries: 4, underwriting: 4, percent: 4, congress: 3

=== 275174newsML.txt ===

local: 12, access: 11, fees: 9, service: 7, internet: 7, companies: 6, providers: 6, fcc: 6, said: 5, phone: 4

=== 277117newsML.txt ===

commission: 8, internet: 8, said: 8, service: 6, notice: 6, access: 6, fees: 5, companies: 5, providers: 4, local: 4

=== 277513newsML.txt ===

rules: 11, key: 9, recovery: 8, products: 7, software: 7, final: 6, draft: 5, would: 5, keys: 5, commerce: 4

=== 290125newsML.txt ===

bill: 12, leach: 8, would: 8, banking: 7, said: 7, banks: 5, bank: 5, committee: 4, insurance: 4, companies: 4

=== 299375newsML.txt ===

computer: 6, export: 6, disk: 4, court: 4, limits: 4, karn: 4, code: 4, encryption: 3, speech: 3, bass: 3

=== 312178newsML.txt ===

credit: 15, unions: 10, case: 10, court: 9, union: 5, said: 5, federal: 4, supreme: 4, justices: 4, banks: 4

=== 324896newsML.txt ===

internet: 18, phone: 14, said: 12, companies: 9, study: 6, lines: 6, company: 5, local: 4, system: 4, coalition: 4

=== 325347newsML.txt ===

internet: 18, phone: 14, said: 14, companies: 9, study: 6, lines: 6, company: 5, local: 4, system: 4, access: 4

=== 330967newsML.txt ===

financial: 11, said: 7, banking: 5, administration: 5, would: 5, issue: 4, bill: 4, services: 3, proposal: 3, banks: 3

=== 331411newsML.txt ===

financial: 7, plan: 6, said: 6, companies: 5, banking: 5, would: 5, treasury: 4, bank: 4, allowing: 4, administration: 3

=== 347226newsML.txt ===

companies: 7, export: 7, encryption: 6, stronger: 5, programmes: 5, said: 5, computer: 3, commerce: 3, information: 3, software: 3

=== 354135newsML.txt ===

said: 10, encryption: 9, draft: 9, guidelines: 8, policy: 7, group: 5, countries: 5, products: 4, law: 4, dryden: 4

=== 354285newsML.txt ===

draft: 9, encryption: 8, policy: 8, guidelines: 7, said: 7, countries: 5, law: 4, dryden: 4, global: 3, group: 3

=== 357147newsML.txt ===

internet: 14, plan: 10, addresses: 6, address: 6, new: 5, domains: 5, said: 5, disputes: 4, would: 4, registration: 4

=== 366020newsML.txt ===

financial: 7, banking: 7, restrictions: 6, leach: 5, reform: 4, insurance: 4, companies: 4, said: 4, federal: 4, industry: 3

=== 369570newsML.txt ===

banks: 9, financial: 8, insurance: 8, banking: 7, congress: 5, companies: 5, said: 5, services: 4, reform: 4, introduced: 4

=== 371380newsML.txt ===

internet: 10, said: 7, congress: 7, legislators: 5, also: 4, president: 4, caucus: 3, laws: 3, would: 3, providers: 2

=== 372744newsML.txt ===

bill: 10, said: 9, administration: 7, export: 4, goodlatte: 4, would: 4, burns: 4, encryption: 3, technology: 3, last: 3

=== 372989newsML.txt ===

internet: 10, congress: 8, said: 7, legislators: 5, also: 4, president: 4, would: 4, caucus: 3, laws: 3, encryption: 3

=== 372995newsML.txt ===

internet: 10, said: 7, congress: 7, legislators: 5, also: 4, president: 4, caucus: 3, laws: 3, would: 3, providers: 2

=== 378457newsML.txt ===

insurance: 12, banks: 9, said: 9, rhode: 5, island: 5, law: 5, states: 5, state: 4, national: 4, selling: 4

=== 394237newsML.txt ===

said: 13, greenspan: 9, federal: 6, congress: 6, financial: 6, regulators: 6, fed: 6, trading: 6, bank: 5, would: 5

=== 398094newsML.txt ===

credit: 20, unions: 12, court: 11, banks: 9, union: 9, said: 8, federal: 7, supreme: 4, members: 4, decision: 3

=== 401260newsML.txt ===

foreign: 15, cftc: 13, court: 11, trading: 10, exchange: 10, currency: 9, amendment: 7, supreme: 6, futures: 5, commission: 5

=== 407599newsML.txt ===

bills: 8, encryption: 7, key: 7, said: 7, export: 6, products: 5, would: 5, administration: 4, clinton: 3, computer: 3

Stylistic Analysis

Total unique words: 3304

Total word count: 25834

Type Token Ratio (TTR): 12.79%

Total stop-words count: 8689

Most common Parts of Speech n-grams:

1. ('DT', 'NN'), 1451

2. ('JJ', 'NN'), 1138

3. ('IN', 'DT'), 1015

4. ('NN', 'IN'), 996

5. ('NN', 'NN'), 950

6. ('NN', ','), 702

7. ('DT', 'JJ'), 624

8. ('NN', 'NNS'), 557

9. ('IN', 'NN'), 534

10. ('JJ', 'NNS'), 534

11. ('TO', 'VB'), 460
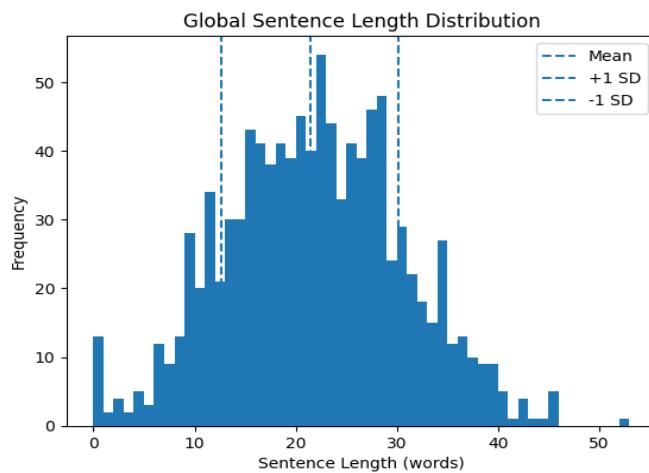
12. ('NN', '.'), 438

13. ('NNS', 'IN'), 420
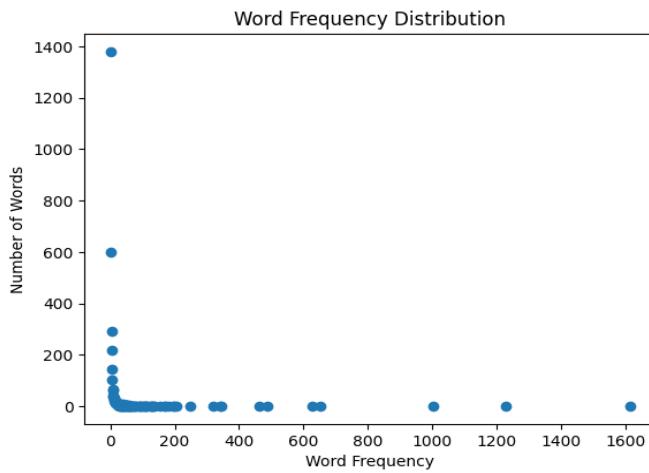
14. ('NN', 'VBD'), 394

15. ('IN', 'JJ'), 348
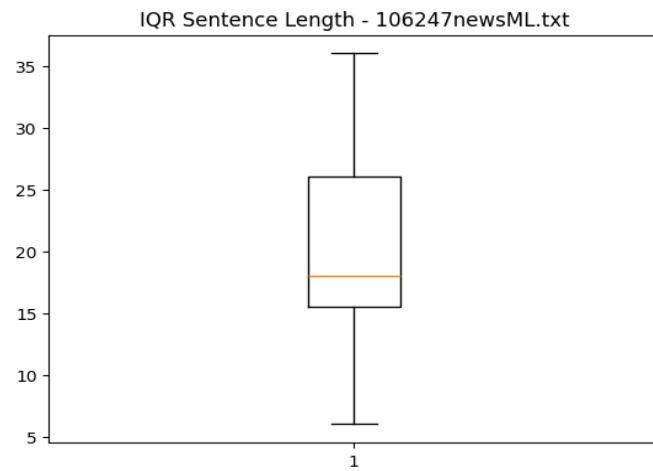
global_sentence_length_distribution.png
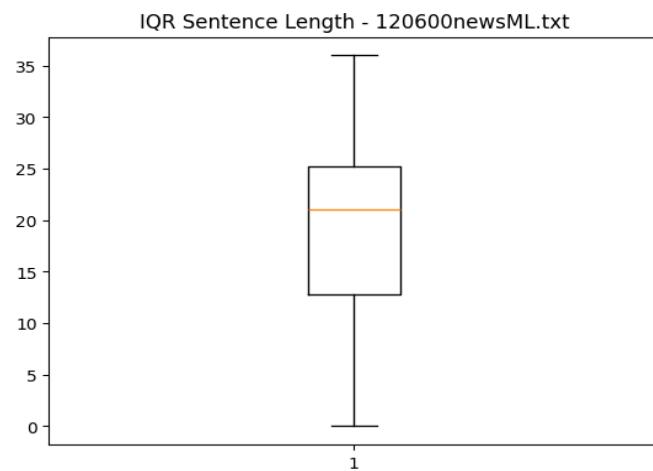
global_word_freq_distribution.png
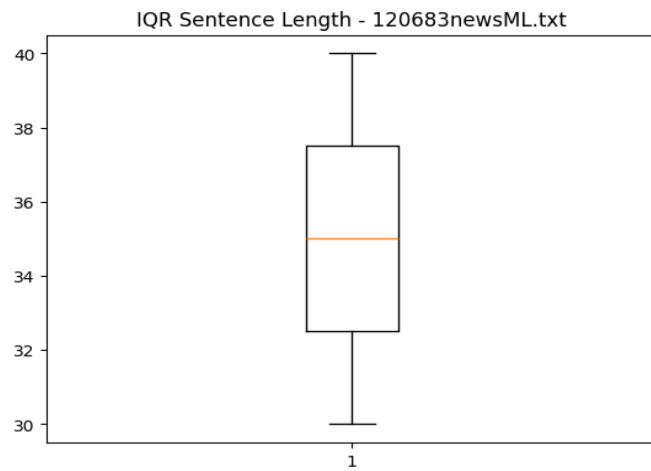
sentence_length_IQR_106247newsML.txt.png

IQR Sentence Length - 106247newsML.txt
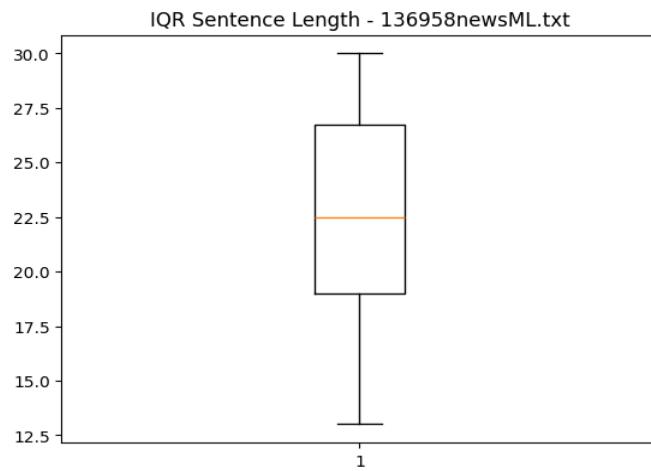
sentence_length_IQR_120600newsML.txt.png

IQR Sentence Length - 120600newsML.txt
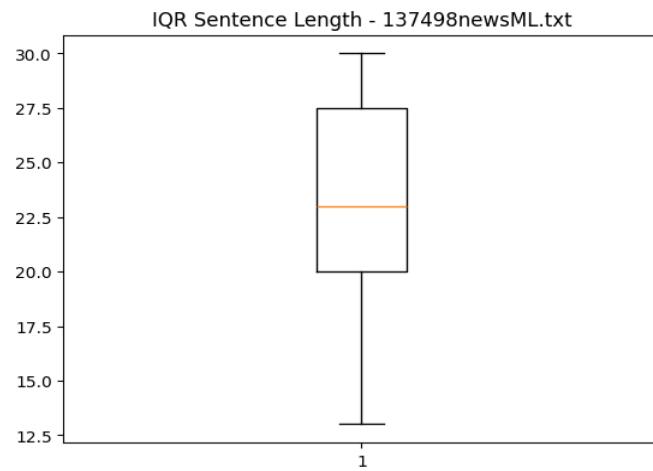
sentence_length_IQR_120683newsML.txt.png

IQR Sentence Length - 120683newsML.txt
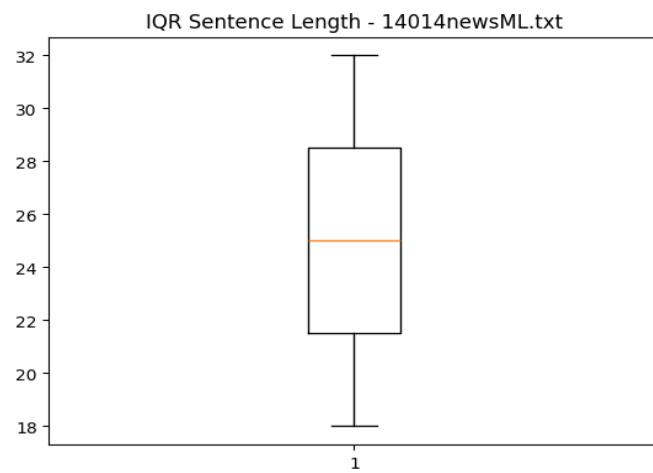
sentence_length_IQR_136958newsML.txt.png

IQR Sentence Length - 136958newsML.txt
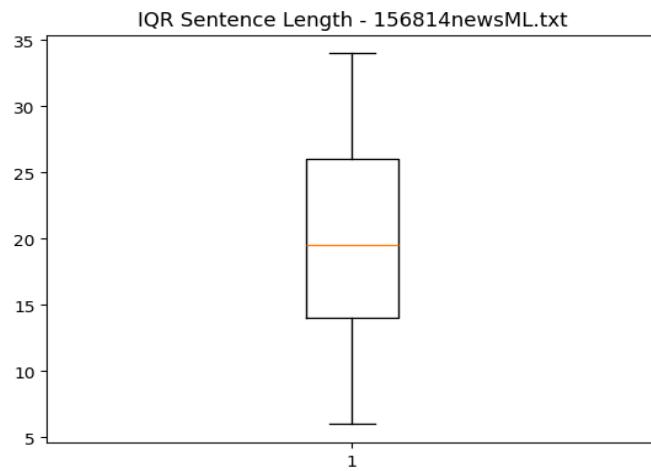
sentence_length_IQR_137498newsML.txt.png

IQR Sentence Length - 137498newsML.txt
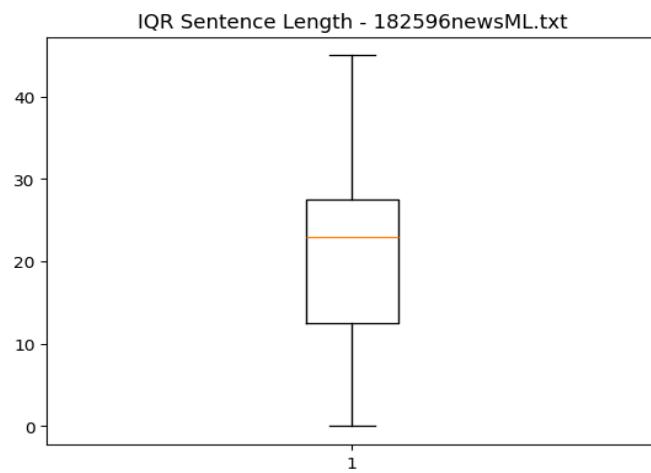
sentence_length_IQR_14014newsML.txt.png

IQR Sentence Length - 14014newsML.txt
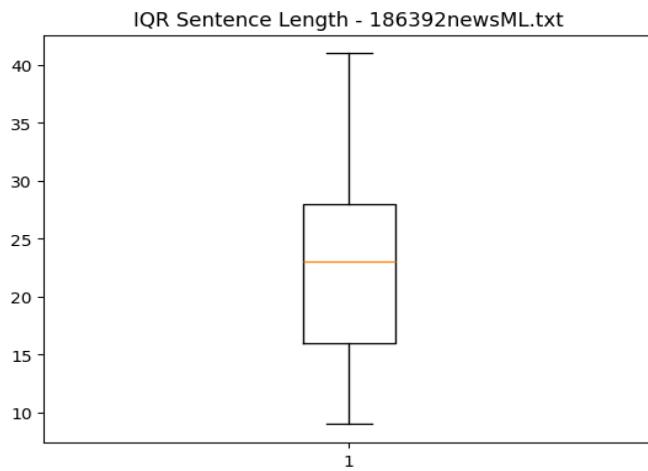
sentence_length_IQR_156814newsML.txt.png

IQR Sentence Length - 156814newsML.txt
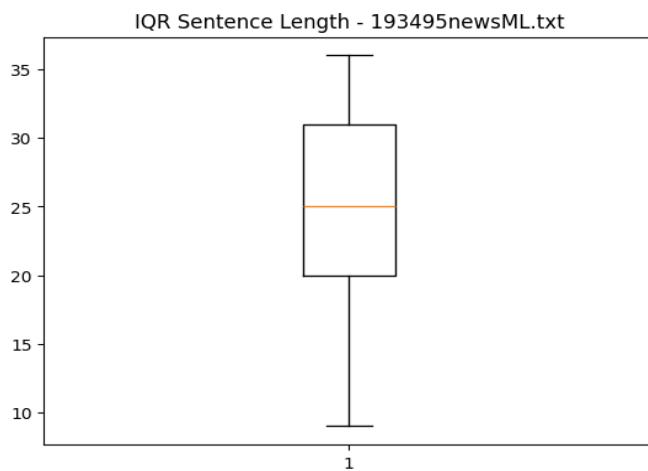
sentence_length_IQR_182596newsML.txt.png

IQR Sentence Length - 182596newsML.txt
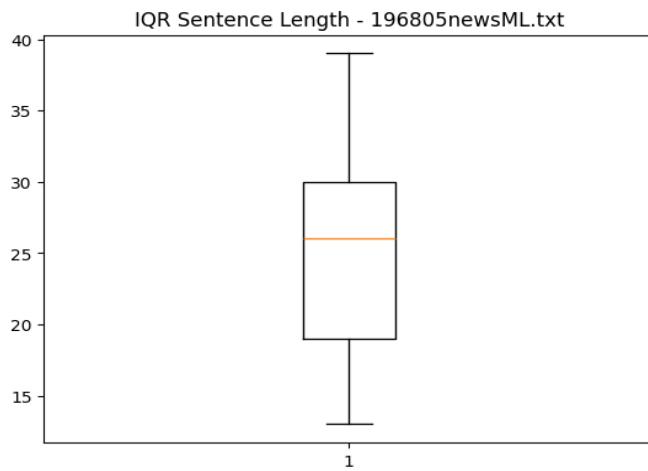
sentence_length_IQR_186392newsML.txt.png

IQR Sentence Length - 186392newsML.txt
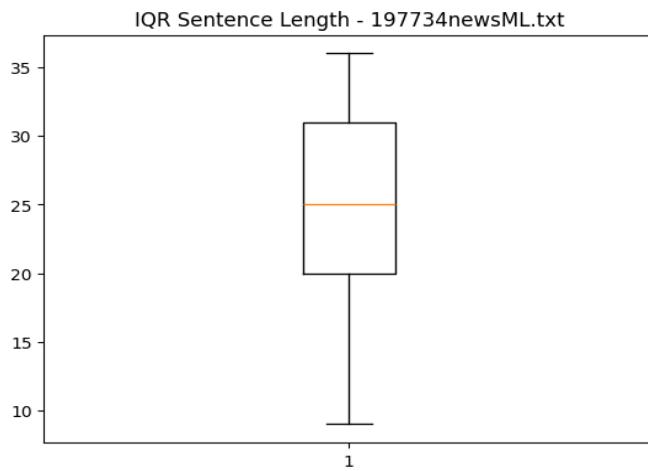
sentence_length_IQR_193495newsML.txt.png

IQR Sentence Length - 193495newsML.txt
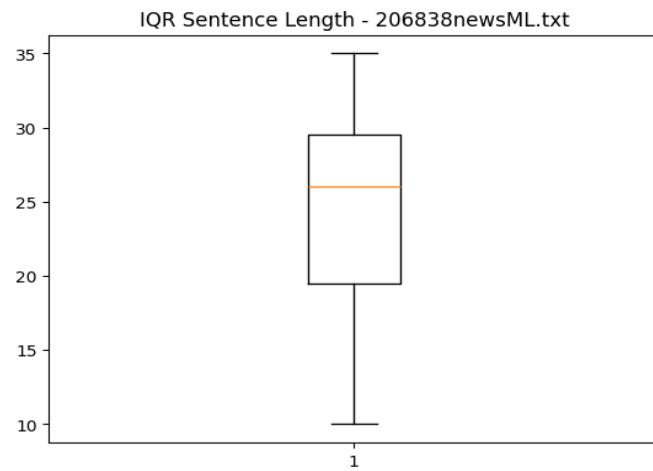
sentence_length_IQR_196805newsML.txt.png

IQR Sentence Length - 196805newsML.txt
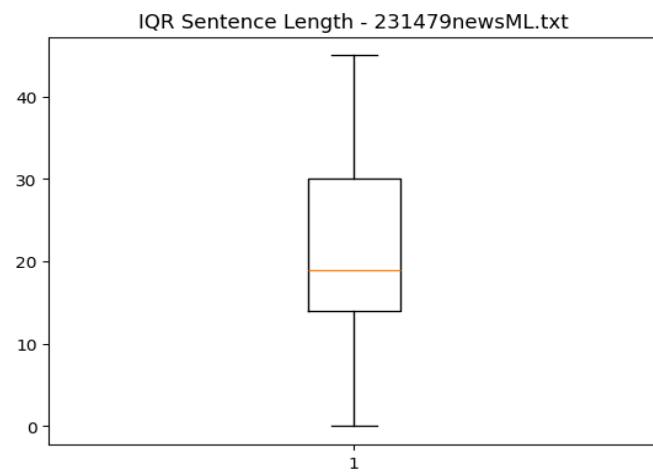
sentence_length_IQR_197734newsML.txt.png

IQR Sentence Length - 197734newsML.txt
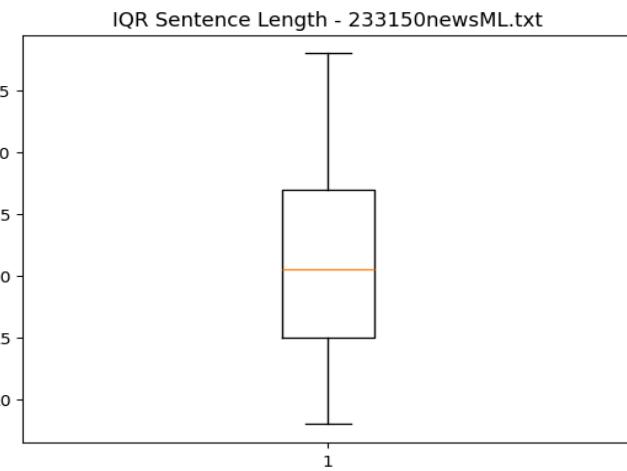
sentence_length_IQR_206838newsML.txt.png

IQR Sentence Length - 206838newsML.txt
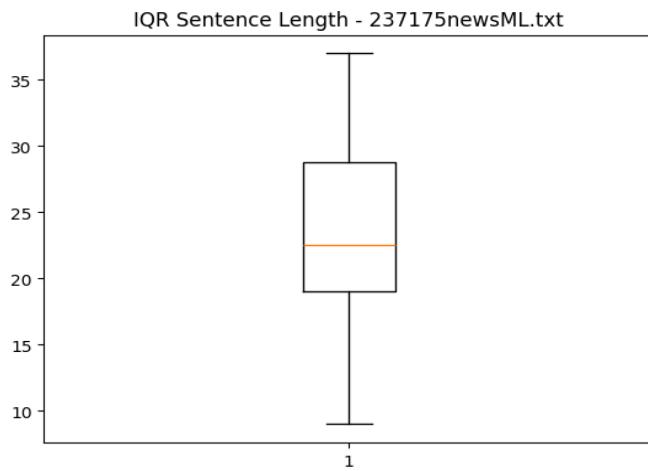
sentence_length_IQR_231479newsML.txt.png

IQR Sentence Length - 231479newsML.txt
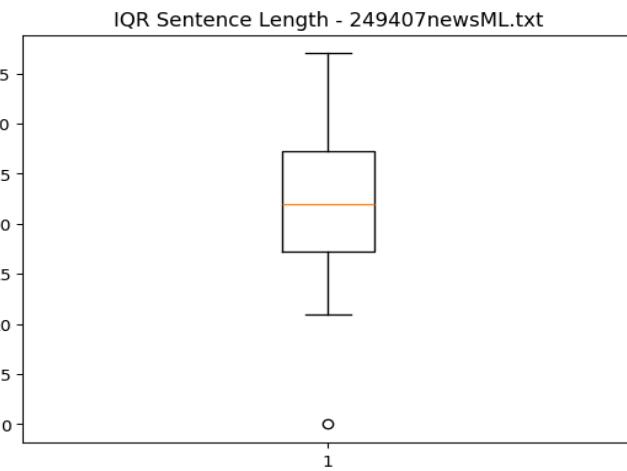
sentence_length_IQR_233150newsML.txt.png

IQR Sentence Length - 233150newsML.txt
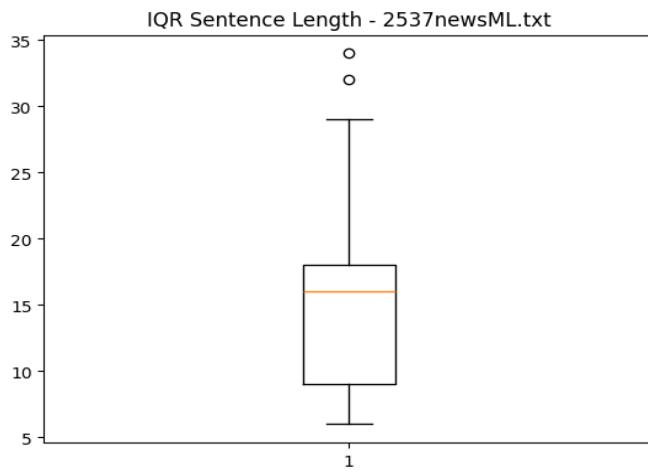
sentence_length_IQR_237175newsML.txt.png

IQR Sentence Length - 237175newsML.txt
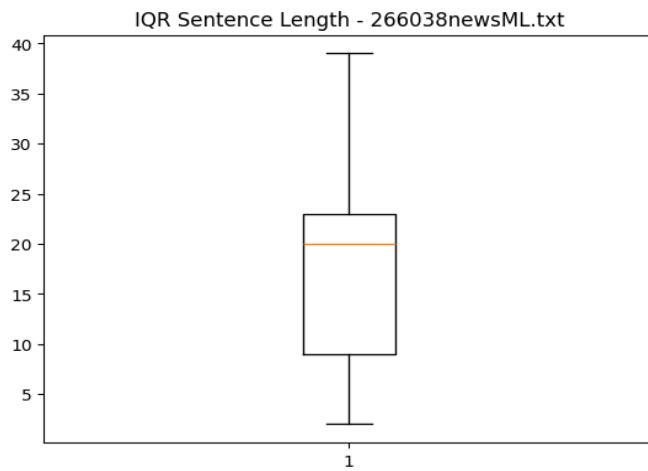
sentence_length_IQR_249407newsML.txt.png

IQR Sentence Length - 249407newsML.txt
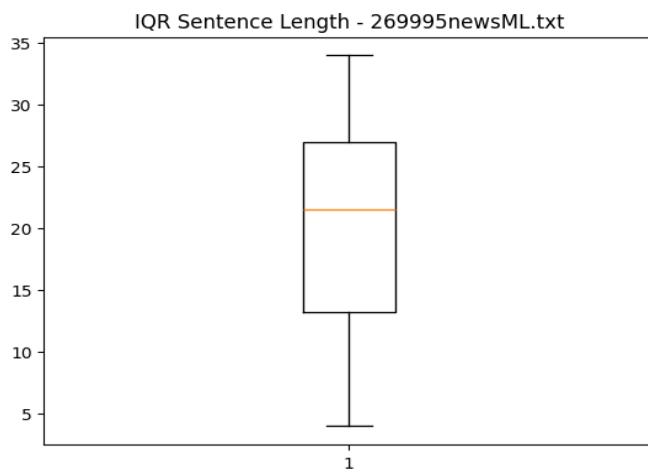
sentence_length_IQR_2537newsML.txt.png

IQR Sentence Length - 2537newsML.txt
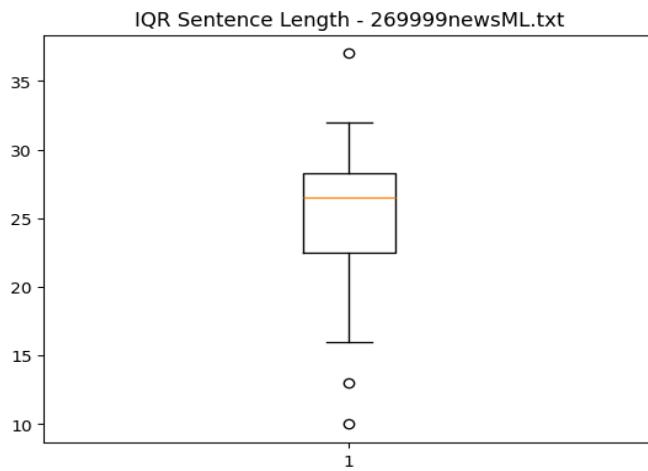
sentence_length_IQR_266038newsML.txt.png

IQR Sentence Length - 266038newsML.txt
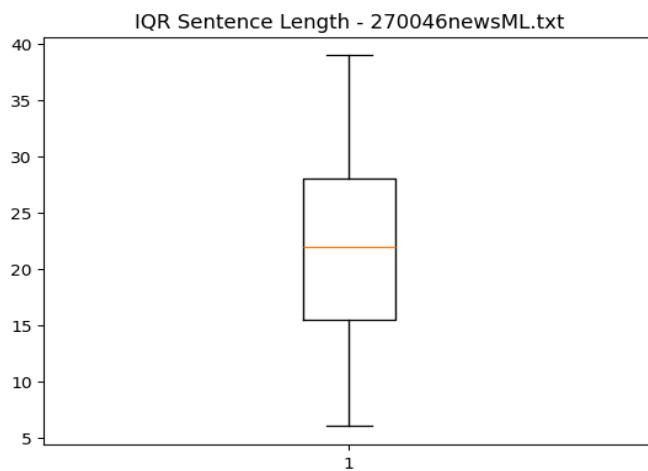
sentence_length_IQR_269995newsML.txt.png

IQR Sentence Length - 269995newsML.txt
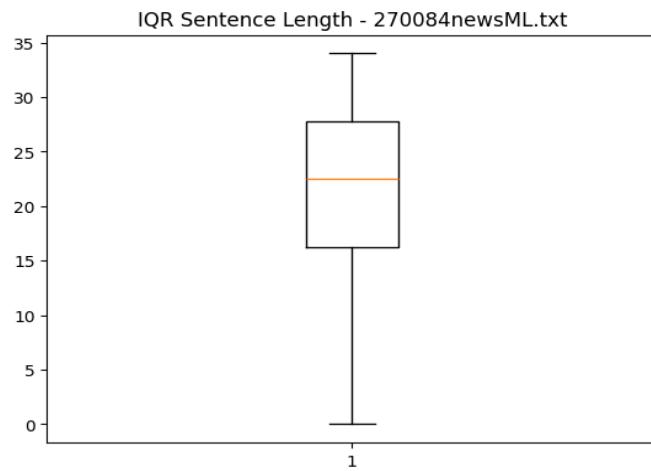
sentence_length_IQR_269999newsML.txt.png

IQR Sentence Length - 269999newsML.txt
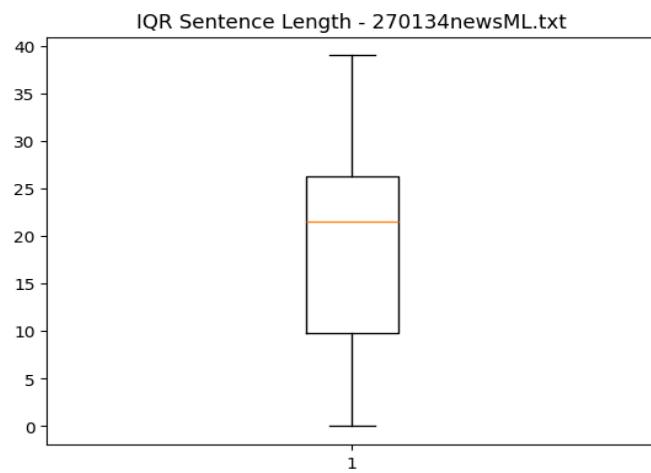
sentence_length_IQR_270046newsML.txt.png

IQR Sentence Length - 270046newsML.txt
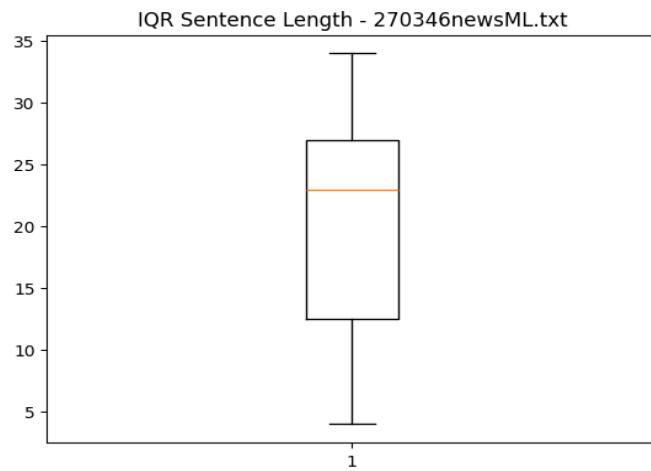
sentence_length_IQR_270084newsML.txt.png

IQR Sentence Length - 270084newsML.txt
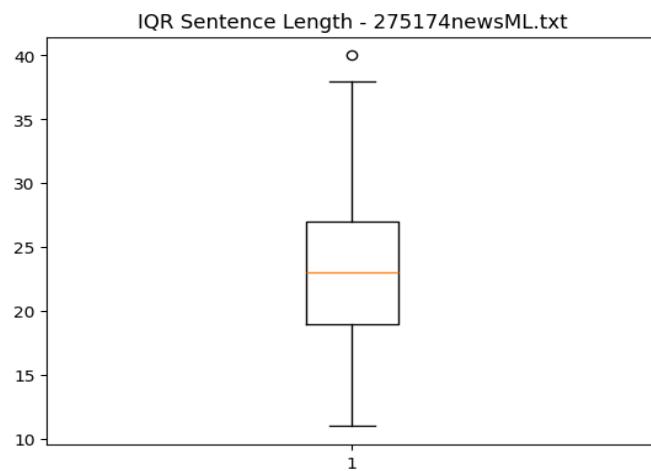
sentence_length_IQR_270134newsML.txt.png

IQR Sentence Length - 270134newsML.txt
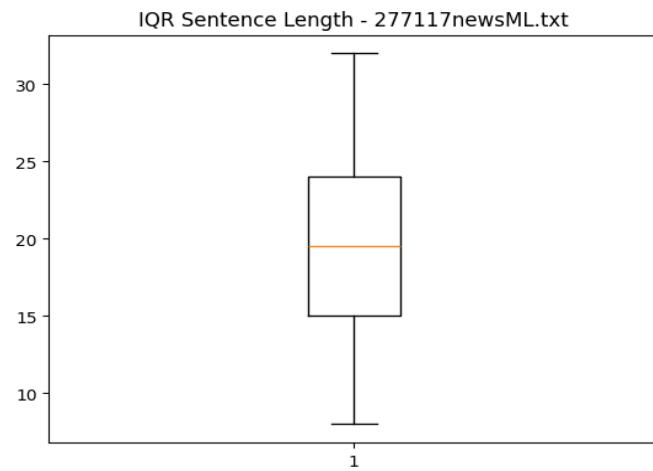
sentence_length_IQR_270346newsML.txt.png

IQR Sentence Length - 270346newsML.txt
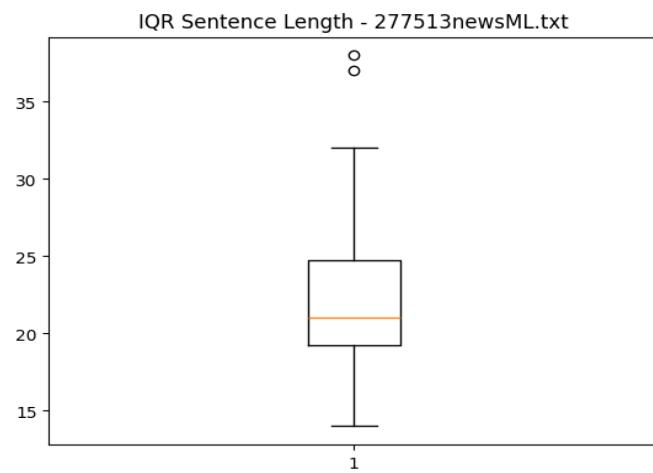
sentence_length_IQR_275174newsML.txt.png

IQR Sentence Length - 275174newsML.txt
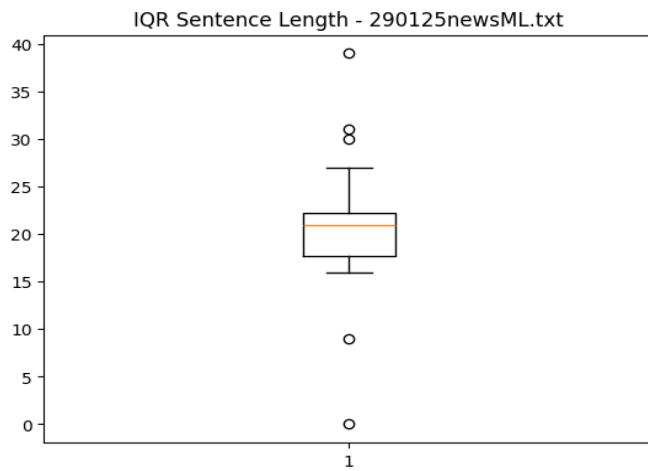
sentence_length_IQR_277117newsML.txt.png

IQR Sentence Length - 277117newsML.txt
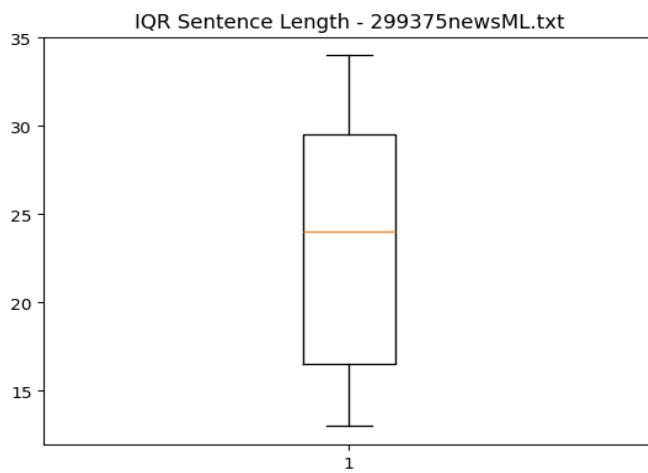
sentence_length_IQR_277513newsML.txt.png

IQR Sentence Length - 277513newsML.txt
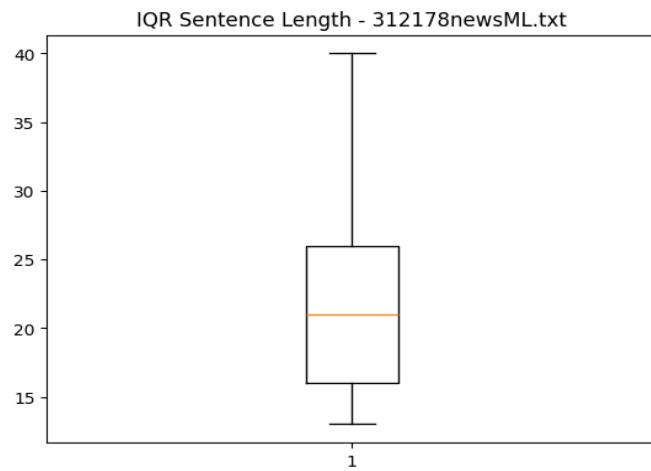
sentence_length_IQR_290125newsML.txt.png

IQR Sentence Length - 290125newsML.txt
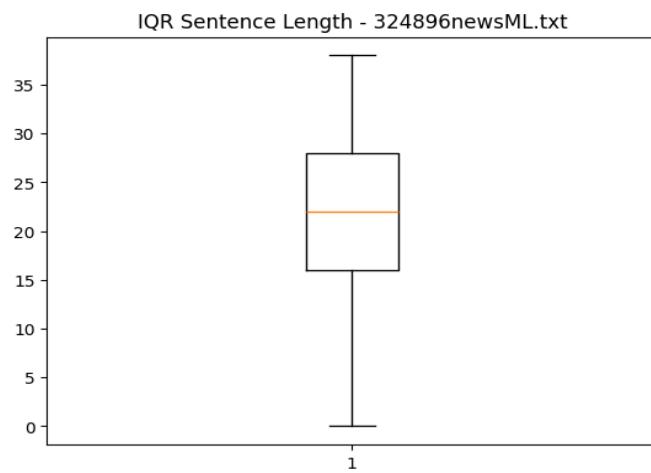
sentence_length_IQR_299375newsML.txt.png

IQR Sentence Length - 299375newsML.txt
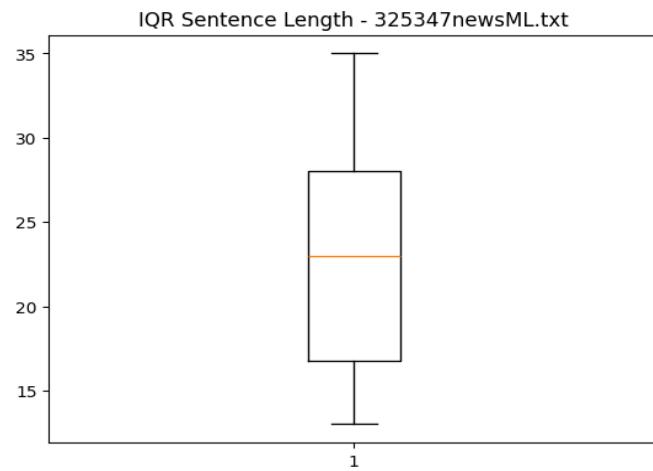
sentence_length_IQR_312178newsML.txt.png

IQR Sentence Length - 312178newsML.txt
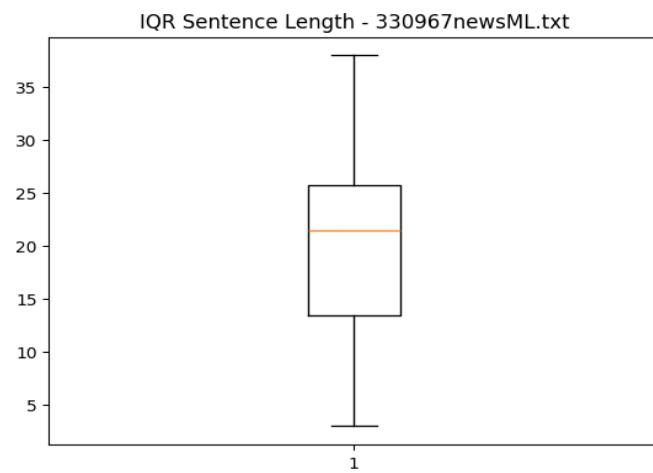
sentence_length_IQR_324896newsML.txt.png

IQR Sentence Length - 324896newsML.txt
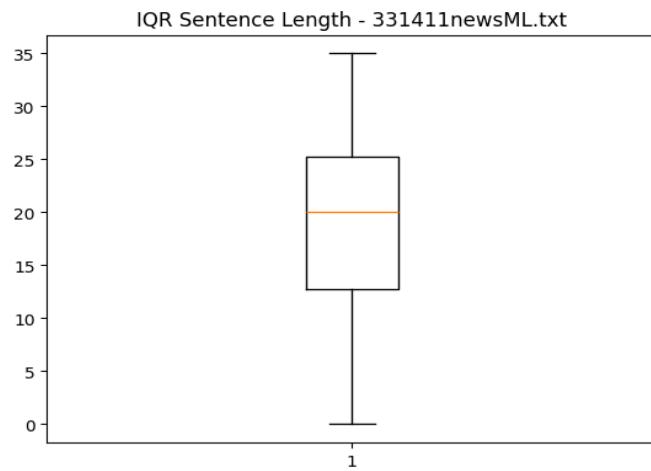
sentence_length_IQR_325347newsML.txt.png

IQR Sentence Length - 325347newsML.txt
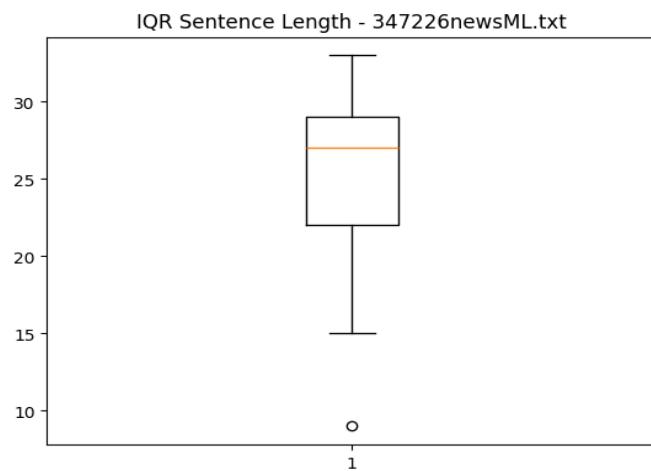
sentence_length_IQR_330967newsML.txt.png

IQR Sentence Length - 330967newsML.txt
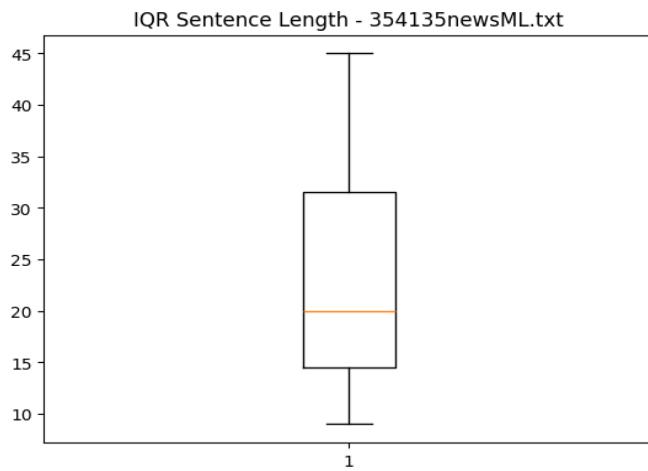
sentence_length_IQR_331411newsML.txt.png

IQR Sentence Length - 331411newsML.txt
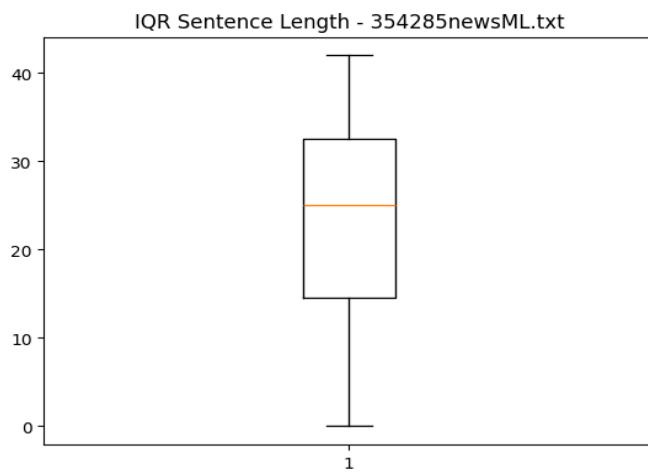
sentence_length_IQR_347226newsML.txt.png

IQR Sentence Length - 347226newsML.txt
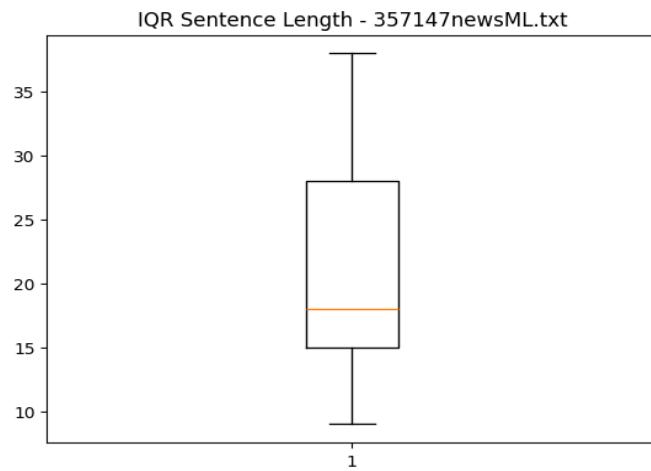
sentence_length_IQR_354135newsML.txt.png

IQR Sentence Length - 354135newsML.txt
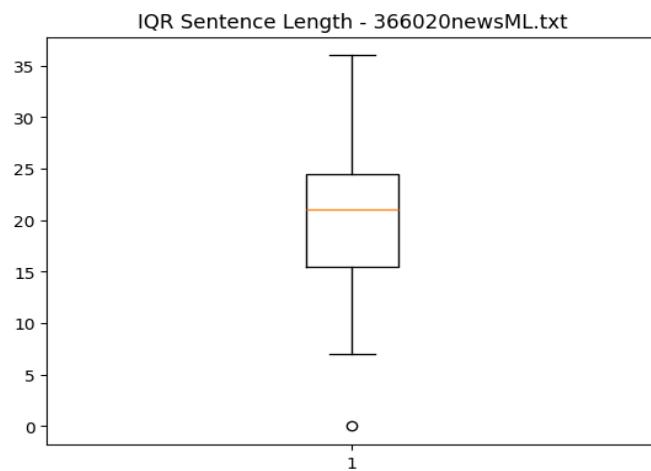
sentence_length_IQR_354285newsML.txt.png

IQR Sentence Length - 354285newsML.txt
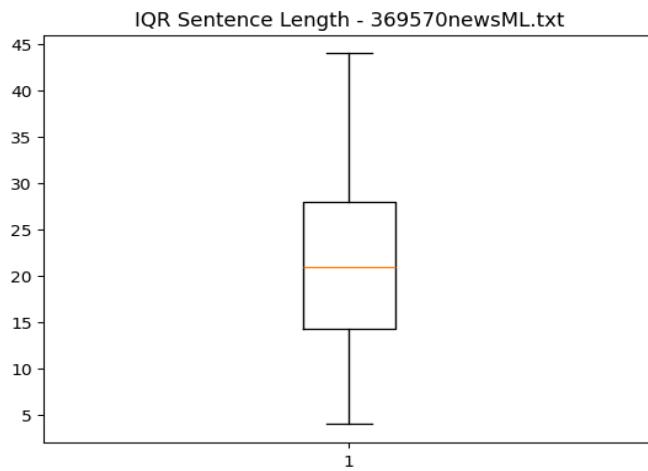
sentence_length_IQR_357147newsML.txt.png

IQR Sentence Length - 357147newsML.txt
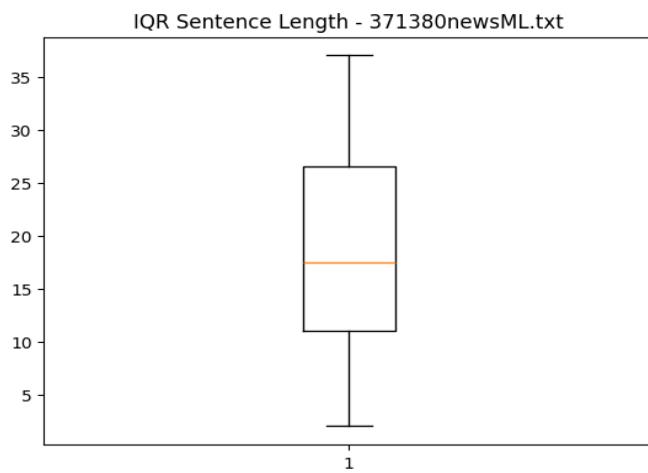
sentence_length_IQR_366020newsML.txt.png

IQR Sentence Length - 366020newsML.txt
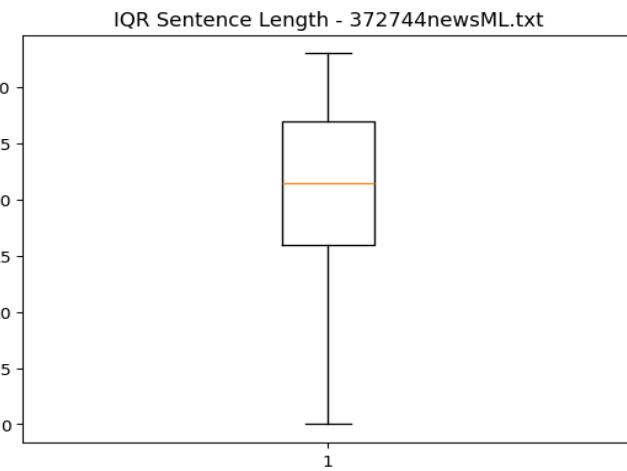
sentence_length_IQR_369570newsML.txt.png


IQR Sentence Length - 369570newsML.txt
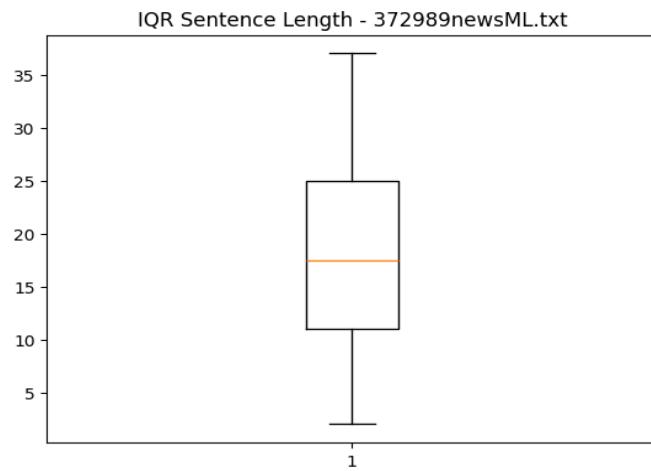
sentence_length_IQR_371380newsML.txt.png

IQR Sentence Length - 371380newsML.txt

sentence_length_IQR_372744newsML.txt.png

IQR Sentence Length - 372744newsML.txt
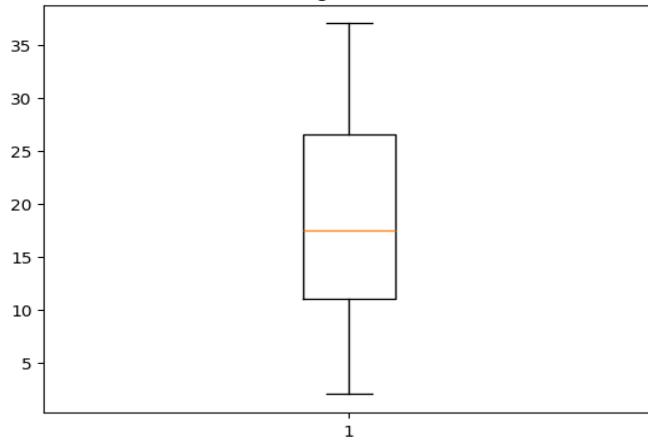
sentence_length_IQR_372989newsML.txt.png

IQR Sentence Length - 372989newsML.txt
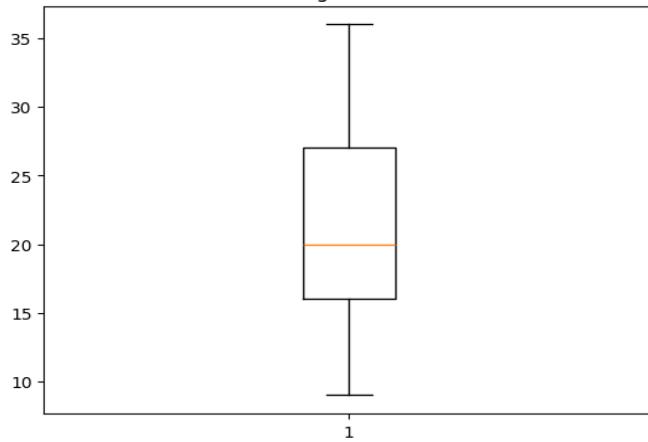
sentence_length_IQR_372995newsML.txt.png

IQR Sentence Length - 372995newsML.txt
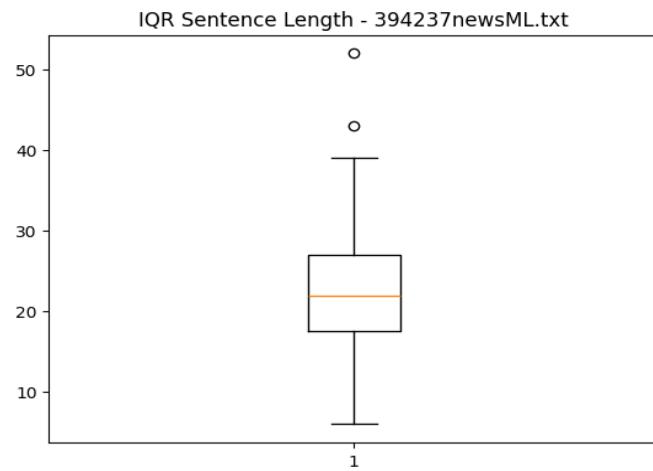
sentence_length_IQR_378457newsML.txt.png
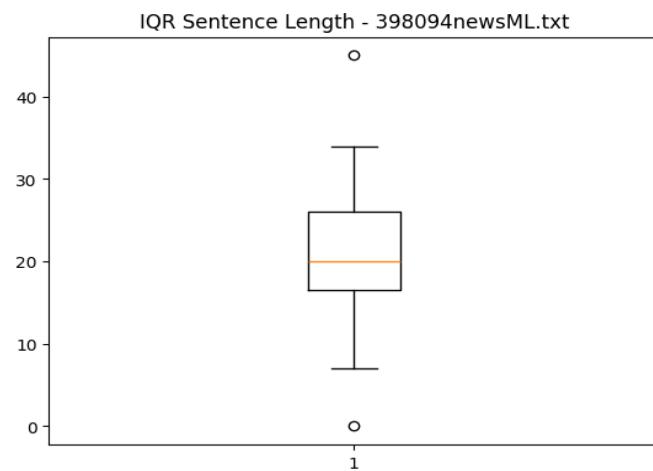
IQR Sentence Length - 378457newsML.txt
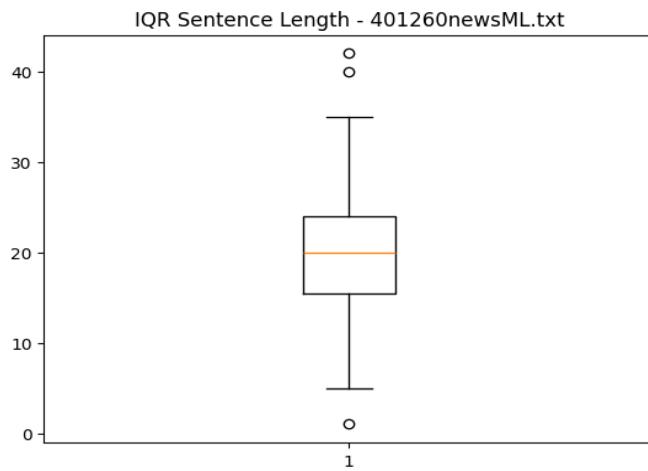
sentence_length_IQR_394237newsML.txt.png

IQR Sentence Length - 394237newsML.txt

sentence_length_IQR_398094newsML.txt.png

IQR Sentence Length - 398094newsML.txt

sentence_length_IQR_401260newsML.txt.png

IQR Sentence Length - 401260newsML.txt

sentence_length_IQR_407599newsML.txt.png

IQR Sentence Length - 407599newsML.txt